AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL075502

Key Points:
- A coupled hydrologic/hydrodynamic model cascade evaluates flood inundation and volume simulated using reanalysis meteorological data
- We used Australia as a case study to demonstrate the feasibility of a continental implementation for continuous flood event simulations
- Reanalysis simulations capture flood inundated areas and volumes although the meteorological ensemble spread results in large uncertainty

Supporting Information:
- Supporting Information S1

Correspondence to:
K. M. Andreadis,
kandread@jpl.nasa.gov

Citation:
Andreadis, K. M., Schumann, G. J.-P., Stampoulis, D., Bates, P. D., Brakenridge, G. R., & Kettner, A. J. (2017). Can atmospheric reanalysis datasets be used to reproduce flooding over large scales? *Geophysical Research Letters*, *44*. https://doi.org/10.1002/2017GL075502

Received 5 SEP 2017
Accepted 27 SEP 2017
Accepted article online 4 OCT 2017

Can Atmospheric Reanalysis Data Sets Be Used to Reproduce Flooding Over Large Scales?

Konstantinos M. Andreadis[1], Guy J.-P. Schumann[2,3], Dimitrios Stampoulis[4], Paul D. Bates[3], G. Robert Brakenridge[5], and Albert J. Kettner[5]

[1] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [2] Remote Sensing Solutions, Inc., Monrovia, CA, USA, [3] School of Geographical Sciences, University of Bristol, Bristol, UK, [4] Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, USA, [5] Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, CO, USA

Abstract Floods are costly to global economies and can be exceptionally lethal. The ability to produce consistent flood hazard maps over large areas could provide a significant contribution to reducing such losses, as the lack of knowledge concerning flood risk is a major factor in the transformation of river floods into flood disasters. In order to accurately reproduce flooding in river channels and floodplains, high spatial resolution hydrodynamic models are needed. Despite being computationally expensive, recent advances have made their continental to global implementation feasible, although inputs for long-term simulations may require the use of reanalysis meteorological products especially in data-poor regions. We employ a coupled hydrologic/hydrodynamic model cascade forced by the 20CRv2 reanalysis data set and evaluate its ability to reproduce flood inundation area and volume for Australia during the 1973–2012 period. Ensemble simulations using the reanalysis data were performed to account for uncertainty in the meteorology and compared with a validated benchmark simulation. Results show that the reanalysis ensemble capture the inundated areas and volumes relatively well, with correlations for the ensemble mean of 0.82 and 0.85 for area and volume, respectively, although the meteorological ensemble spread propagates in large uncertainty of the simulated flood characteristics.

1. Introduction

Floods are one of the most devastating natural disasters, affected an estimated 2.8 billion people in the past 30 years (Doocy et al., 2013) and causing damages with costs of U.S.$6 billion annually (Hallegatte et al., 2013). Population exposure varies by nation: 15 nations together account for 80% of population exposed to river flood risk worldwide (Winsemius et al., 2013). Flood hazard maps have long been prepared in the developed nations at the local scale, but costs are high for such detailed hydraulic modeling as it must be based on abundant in situ data about floodplain topography and channel bathymetry. Such work requires in any case abundant hydrological data, such as river flow time series, which are not available for many of the nations most affected by damaging floods. Thus, for many regions across the globe, a different approach is needed (Hagen & Lu, 2011). Efforts to estimate flood risk at continental and global scales have increased recently, either through remote sensing observations (Brakenridge et al., 2017) or models (Ward et al., 2015). Satellite observations can be hindered by sparse spatial and temporal coverage (due to cloud cover or orbital characteristics), and the inability to only map flooded area but not depths (directly linked to flood hazard) in the case of the more prevalent optical sensors. Global flood risk models can overcome those limitations, although they have been restricted to only estimating river discharge and either solving floodplain dynamics at inadequate spatial resolutions (tens of kilometers) (e.g., Yamazaki et al., 2011) or using simplified physics (e.g., Paiva et al., 2013).

Recently, advances in numerical algorithms and high-performance computing as well as the availability of global data sets (e.g., topography) have allowed the implementation of hydrodynamic models over continents and globally (e.g., Sampson et al., 2015). The fidelity of these models, and particularly their spatial resolution (1 km or less), has enabled them to resolve some of the fine-scale processes that control flooding. Nevertheless, most of the existing global model applications have focused on the estimation of flood characteristics for different flow return periods (Trigg et al., 2016), instead of simulating event-continuous time series of

flood inundation. The transformation of streamflow to flood inundation can be highly nonlinear, which can lead to noncorrespondence between their probability distributions. The latter has implications for flood risk assessment, and it has been posited that in order to correctly assign flood return periods, long-term records of streamflow need to be supplemented by floodplain inundation and depth (Schumann et al., 2016).

Most applications that use return periods assume the flow with the same probability occurs everywhere simultaneously. This problem may be alleviated by obtaining boundary inflows from in situ measurements, using sophisticated statistical models to produce realistic flood event footprints from flood frequency distributions (Dixon & Tawn, 1995), or by forcing hydrodynamic models with runoff generated from meteorological fields that produce realistic flood events. The latter choice essentially creates a model cascade of loosely coupled models that begins with the meteorological data that force the hydrologic model, which in turn produces the streamflow data that are used as boundary conditions for the hydrodynamic model resulting in maps of flood-pertinent variables (Pappenberger et al., 2012). In order to produce historical, long-term flood event time series from such a model cascade we need consistent meteorology data sets that are available globally. Reanalysis data sets, which are produced by assimilating multisensor observations into a climate model, could be a viable option for reconstructing flood events globally especially in data-poor regions (e.g., developing countries).

Previous studies have evaluated whether meteorology derived from reanalysis data sets can reproduce the hydrology of river basins at multiple spatial and temporal scales (e.g., Essou et al., 2016). In addition, such studies within the context of flood modeling have either focused on river flow (e.g., Zsótér et al., 2016), have been performed at relatively coarse spatial scales (e.g., Emerton et al., 2017), or have only simulated flood return periods but not event time series (e.g., Ward et al., 2013). However, there has been relatively little work assessing whether such data sets can simulate flood inundation via a model cascade and capture the variability of flood events at scales that adequately resolve floodplain hydrodynamics. Although uncertainties in meteorological data, such as the ones derived from reanalysis products, may be too large for accurate local-scale studies (Sampson et al., 2014) there is a need to evaluate the skill of reanalysis data over appropriate scales.

Here we evaluate the ability of an atmospheric reanalysis data set, when used as input to a coupled hydrology/hydrodynamic model, to reproduce flood inundation of a continental-size area over a 40 year period. Although the overarching goal of the reanalysis-driven reconstruction of flood events is its evaluation globally, we use Australia as a case study and stepping stone toward global implementation. Apart from its size, which would demonstrate the feasibility of long-term hydrodynamic simulations, the availability of a validated, benchmark simulation that would allow for a rigorous evaluation supports Australia as an appropriate test case.

2. Methods

The experimental design involves the prediction of flood characteristics, such as inundated area and volume, from a coupled hydrologic and hydrodynamic model that has been forced by the reanalysis meteorological data sets. A benchmark hydrodynamic simulation (driven by observed streamflow) that has been validated against satellite observations was used to evaluate how well the reanalysis simulations reproduced flood characteristics over Australia.

2.1. Meteorological Data Sets

The meteorological data necessary for driving the hydrologic model are derived from the Twentieth Century Reanalysis (20CRv2) data set (Compo et al., 2011), which is a long-term (1871 up to present) reanalysis product that provides a 56-member ensemble of numerous atmospheric variables at a spatial resolution of 2° (T62) and a time step of 6 h. Here we use daily time series of 20CRv2 precipitation, maximum and minimum 2 m air temperature, and 10 m wind speed in order to match the meteorological forcing requirements of the hydrologic model. When compared with observations, 20CRv2 data have been shown to reproduce the variability of precipitation rates (e.g., Dolinar et al., 2016) while also being used in regional hydrologic studies. For example, Whelan and Frederiksen (2017) utilized 20CRv2 data to examine the role of the Madden-Julian Oscillation in Australian flooding during La Niña events. Although alternative reanalysis data products exist, e.g., Japanese 55-year Reanalysis (Kobayashi et al., 2015), they do not provide uncertainty estimates and would hinder the proper evaluation of how that uncertainty propagates from the reanalysis to the simulated flooded area and volume. In contrast, the 20CRv2 data set provides the mean and spread (i.e., standard deviation) of

the ensemble allowing the stochastic simulation of flood characteristics that could also be directly translated to flood risk.

The reanalysis precipitation was downscaled using the Constructed Analogs (CA) method (Hidalgo et al., 2008), while air temperature and wind speed were spatially downscaled using the methods from Sheffield et al. (2006). The latter involved bilinear interpolation of the reanalysis fields to the model 0.25° grid, with adjustments made to air temperature for elevation effects. Before interpolation to the finer-scale grid, the data were adjusted to sea level using a lapse rate of 6.5°C km^{-1} and readjusted to the model elevation after interpolation. The CA method is a deterministic approach for statistically downscaling coarse-scale meteorological fields to finer spatial scales by linearly combining past weather patterns that exhibit similarity to the coarse-scale (i.e., target) pattern. The past weather patterns (i.e., analogs) used to derive the linear combination (i.e., regression) are obtained by aggregating finer-scale data to the target resolution. The downscaled fields are then constructed by applying the linear regression coefficients for the target pattern to the finer-scale data from the same days used to derive the analog (Hidalgo et al., 2009).

In this study, the analog patterns are composed of daily precipitation observed from the Tropical Rainfall Measuring Mission (TRMM) satellite (Huffman et al., 2007) and aggregated to the spatial resolution of the 20CRv2 data set. The TRMM data used here (3b42v7 product) cover the period 1 January 1998 to 31 December 2012 and have a spatial resolution of 0.25°. The target pattern was estimated by performing a regression on the 30 best predictors ranked in terms of the coarse-scale spatial root-mean-square error, with the potential patterns being selected within ±45 days from the target day (Maurer & Hidalgo, 2008). The resulting regression coefficients for each target day were then applied to the fine-scale patterns (from TRMM) from the same days as the selected predictors. Given the objective of this study, which is the evaluation of reanalysis data sets for reproducing flood inundation, we opted to not perform any bias correction (e.g., Maurer et al., 2010) so as not to modify the magnitude or timing of the reanalysis precipitation. In addition, despite the availability of a long-term ground-based data set of precipitation over Australia (Jones et al., 2009), which could have been used instead of TRMM, we elected to demonstrate the feasibility of the approach globally where such detailed data would not exist. Although a direct pixel-by-pixel comparison would not be applicable, the resulting spatial patterns between the downscaled reanalysis precipitation and the TRMM observations showed good agreement with pattern correlations ranging from 0.81 to 0.94 (maps of seasonal means are shown in Figure S1 in the supporting information).

2.2. Model Description

Flood inundation area and volume were simulated using a state-of-the-art hydrodynamic model, LISFLOOD-FP, which solves an approximation of the 2-D Saint-Venant equation in a computationally efficient manner (Bates et al., 2010). The model is able to not only simulate river and floodplain hydraulics over large areas but also resolve processes for rivers with widths smaller than the model's nominal resolution by using a subgrid channel formulation (Neal et al., 2012). LISFLOOD-FP requires inputs on floodplain topography, river channel widths, and bank elevation, as well as a set of upstream and lateral boundary conditions. Topography was derived at 1 km from Shuttle Radar Topography Mission (SRTM) satellite data, after being corrected for vegetation using a global ICESat canopy data set and hydrologically conditioned (Schumann et al., 2013). River channel widths were derived from a global database (Andreadis et al., 2013), while river bathymetry was estimated within the model based on hydraulic geometry (Neal et al., 2012). Lakes and wetlands were included in the model using the Global Land and Wetlands Database (Lehner & Doll, 2004), and only rivers that drained areas larger than 10,000 km^2 were explicitly modeled. Boundary conditions for the model included inflows upstream of the modeled rivers and at locations with significant flow contributions from tributaries, as well as downstream of rivers where normal depth flow conditions were imposed (shown in Figure S2 in the supporting information along with a map of elevation). A more detailed description of the implementation of LISFLOOD-FP over Australia can be found in Schumann et al. (2016), with hydrodynamic variables being simulated at a 1 km spatial resolution over a period of 40 years (1 January 1973 to 31 December 2012).

The inflows (42 locations) to the LISFLOOD-FP model were obtained from observed daily flows for the benchmark simulation and estimated from a hydrology model for each of the reanalysis simulations. The model used to simulate the upstream and lateral inflows was the Variable Infiltration Capacity (VIC) macroscale model (Liang et al., 1994), which has been used successfully numerous times to simulate the hydrology of large-scale river basins (e.g., Nijssen et al., 2001). The VIC model solves the energy and water balance over a gridded domain, with each grid cell being composed of tiles that have been partitioned based on vegetation cover and

elevation. Apart from land cover and topography, the model also accounts for subgrid variability in soil moisture (by treating capacity distribution probabilistically) and precipitation, while the subsurface is represented as three layers that control the generation of surface runoff and baseflow. The VIC model was implemented at a 0.25° spatial resolution with elevation derived from the SRTM data, vegetation fractions obtained from the Moderate resolution Imaging Spectroradiometer land cover product (Friedl et al., 2010), and soil parameters derived from the Australian Soil Resources Information System database (Johnston et al., 2003). The downscaled precipitation, air temperature, and wind speed for each reanalysis data set were used to force VIC and simulate hydrologic fluxes including runoff and baseflow. The latter was then used as input to a river routing model (Lohmann et al., 1998) to estimate streamflow at each inflow location.

2.3. Model Parameters

The benchmark simulation that we use here does not have any dependence on the reanalysis meteorological (or TRMM) data since the hydrodynamic model is driven by observed flows at the upstream and lateral boundary points. This implementation of LISFLOOD-FP over Australia has been previously calibrated by varying model parameters that control river channel bathymetry, cross-section shape, and roughness (Schumann et al., 2016). The optimal model parameterization was evaluated using Receiver Operating Characteristic curves (shown in Figure S3 in the supporting information) and had an Area Under Curve of 0.81 (with 1.0 signifying perfect performance). Simulated inundated area was validated against observations derived from Landsat optical imagery (Mueller et al., 2016), and resulted in correct predictions (i.e., hit rate) of 89.6% of the domain.

The hydrologic parameters were derived by calibrating the VIC model against the observed streamflow data using an efficient algorithm in order to minimize the computational cost of the necessary simulations. Details of the calibration procedure can be found in the supporting information. In order to isolate the impact of the uncertainties of the reanalysis data set on the simulated flood characteristics, we performed the flood simulations using the calibrated set of model parameters. An ensemble of simulations was generated by forcing the VIC/LISFLOOD-FP model with the ensemble of meteorological data constructed from the 20CRv2 mean and standard deviation. Therefore, any uncertainty in the simulated flood characteristics resulted from the uncertainty in the reanalysis data (the ensemble simulations for each inflow location along with the observed flow are shown in Figure S4 in the supporting information).

3. Results

There have been a number of major flood events in Australia during the 1973–2012 study period. Queensland was severely flooded in 1974 and 2010–2011, with the latter event leading to damages that cost billions of dollars. Additional major flooding occurred in the mid-1970s in eastern Australia, including New South Wales and Victoria with events on the order of 35–100 year recurrence between 1973 and 1978. The 2010–2011 flooding in Queensland was followed by another major event in 2011 in Victoria. New South Wales and Queensland were affected by a flood in the spring of 1990, whereas the northern parts of Australia were flooded during 1998 due to cyclone activity. Although the aforementioned events do not, by any means, form an exhaustive list of flooding in Australia, they can be identified by examining the time series of the gauge-driven (i.e., benchmark) simulation of inundated area and volume. Figure 1 shows the monthly time series of inflow (a), inundation volume (b), outflow (c), and inundated area (d) for the benchmark simulation, the reanalysis ensemble (as propagated from the uncertainty in the 20CRv2 data), and the ensemble mean.

The inflow (Figure 1a), calculated as the sum of the boundary inflows, is generally well captured by the ensemble with only 1.0% and 0.7% of values being underestimated and overestimated, respectively. The uncertainty in the reanalysis precipitation can be rather large, with the ensemble coefficient of variation ranging from 0.15 to 1.41 and a mean value of 0.62, propagating into large uncertainty in the simulated inflow. The respective coefficient of variation for inflow ranges from 0.01 to 1.27 with a mean value of 0.88, which partly explains the observed inflow being captured by the ensemble for 98.3% of the simulation period. The temporal variability is relatively well reproduced with Pearson correlation coefficients between the benchmark and the ensemble mean simulation being 0.75 (0.77 and 0.44 for the upper and lower inflow ensemble bounds, respectively). For continuous event simulation temporal correlation is of greater importance than bias in inflow, since the latter is easier to correct for (particularly when cascading it through a hydrodynamic model) and has been shown to be disadvantageous compared to correlation (Criss & Winston, 2008). In this context, the ensemble mean simulation underestimates the observed peaks while the performance in terms of area and volume is

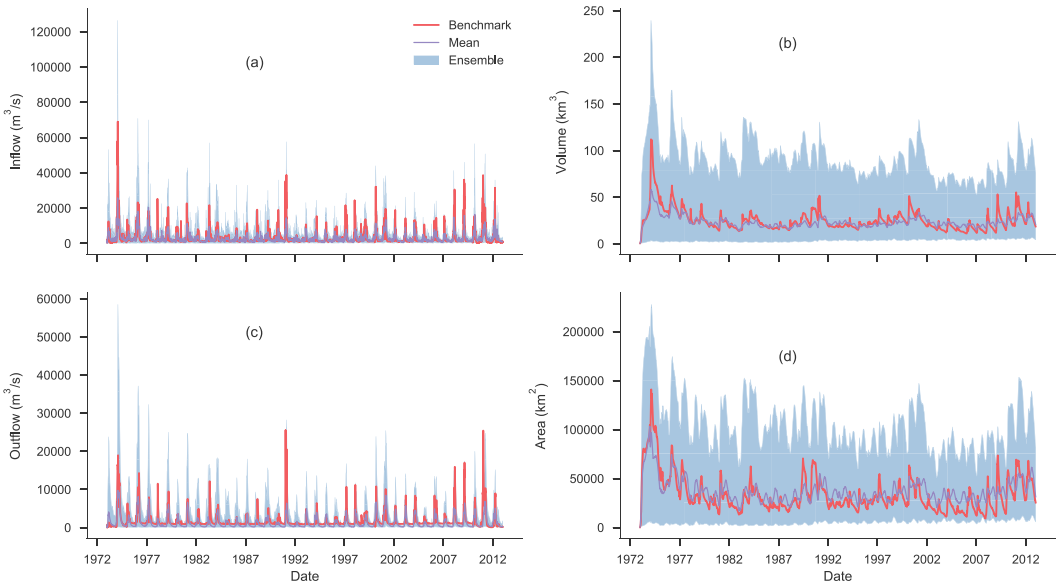

Figure 1. Time series of (a) inflow, (b) inundated volume, (c) outflow, and (d) inundated area from the benchmark and ensemble simulations.

much higher, which can be attributed to the importance of temporal dynamics of successive flood-generating peaks in high-flow periods (Schumann et al., 2016).

The ensemble reanalysis was also able to reproduce the temporal variability in outflows relatively well, with correlations of 0.71, 0.70, and 0.43 for the ensemble mean, maximum, and minimum simulations (Figure 1b). The ensemble mean underestimates the observed outflow by 488.1 m^3/s average during the simulation period. Similar to the simulated inflow, the ensemble spread for outflow is rather large with the average range being 5,237.2 m^3/s. Nonetheless, the ensemble fails to capture and underestimates the observed outflow for 11.2% of the study period. This underestimation suggests that the flow conveyances are miscalculated, and the floodplain waters are not drained back into the river channel correctly.

The upper bound of the ensemble-simulated inflow overestimates the observed flow by 6,188 m^3/s (or 300.1%) on average, leading to a large ensemble spread shown in the inundated volume comparison (Figure 1c). All of the peaks in the inundated volume are captured by the ensemble, with the timing being reproduced relatively well as evidenced by the correlation of 0.85 between the mean and the observations. The ensemble mean generally underestimates the peaks in inundated volume with a bias of -8.78 km^3, while the overall bias (from the daily time series) was -1.67 km^3, suggesting that the mean reanalysis was not able to capture extreme precipitation events well. The spread in the reanalysis data propagates into a relatively large uncertainty in the model predictions of inundated volume, especially during flood events. The ensemble spread of the inundated volume during the annual peak times ranges from 25.0 to 100.8 km^3 (or 134.1 to 173.2% relative to the ensemble mean).

Figure 1d shows the comparison between the reanalysis and the benchmark simulations for inundated area, with results being somewhat similar to the inundated volume comparison. The reanalysis ensemble captures the inundated area from the observation-driven simulation during the entire study period. The correlation between the reanalysis mean and the benchmark simulation is 0.82, suggesting that the timing of overbank flow is being reproduced relatively well. The impact of the meteorological uncertainty is nonexistent in terms of capturing the temporal variability for the ensemble upper bound (correlation of 0.82) but is very large for the lower bound reducing the aforementioned correlation to 0.02 essentially retaining water within the river channels only. Contrary to the inundated volume, the ensemble mean simulation of inundated area has a positive bias when compared with the benchmark simulation ($-3,995.9$ km^2), suggesting that the floodplain residence times (i.e., floodplain drainage) are not accurately reproduced by the reanalysis-driven simulations.

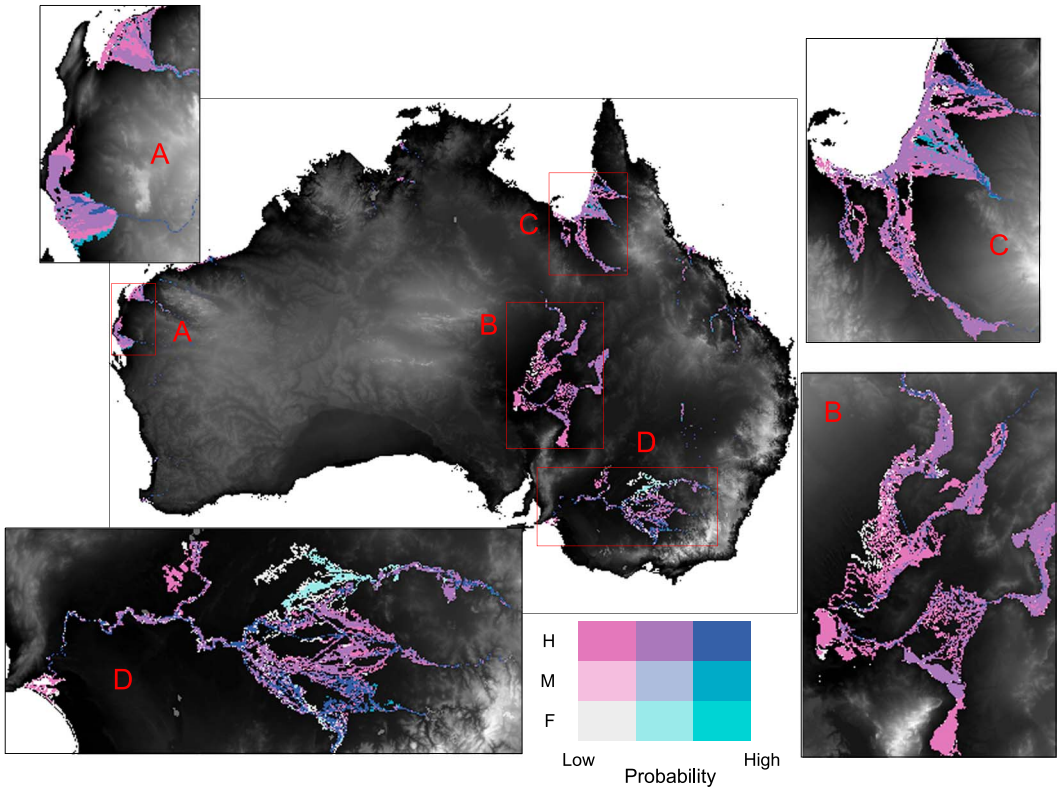

Figure 2. Map of agreement between reanalysis ensemble simulations of maximum inundation area with benchmark simulation over 1973–2012 period. Pixels are categorized in terms of classification metrics (H, Hit; M, Miss; and F, False alarm) and probability of occurrence (i.e., number of ensemble members that had the corresponding classification). Enlarged versions of the inset maps are included in the supporting information as Figures S5–S8.

Although the comparison of the total inundated area and volumes showed that the reanalysis simulations (mean and maximum) mostly captured the temporal variability when compared with the benchmark simulation, we need to also evaluate them in a spatial context. Figure 2 shows a bivariate choropleth map of the maximum inundated area draped over the topography of the study domain. The inundated pixels are colorized according to nine classes derived from two variables with three classes each: a classification metric and the probability of occurrence. The former corresponds to whether the reanalysis simulations predict a pixel as inundated or not in comparison to the benchmark simulation. When both the reanalysis and the benchmark simulation predict a pixel as inundated, it is classified as a Hit (i.e., True positive); when the pixel is inundated in the benchmark but not in the reanalysis simulation, it is classified as a Miss (i.e., False negative); and when the pixel is inundated in the reanalysis but not in the benchmark simulation it is classified as a False alarm (i.e., False positive). Additionally, each pixel is assigned a probability according to the number of ensemble members that agreed with the aforementioned classification, in order to capture the uncertainty in the predictions of inundation extent.

The area that was correctly classified as inundated, in terms of maximum extent, in the reanalysis ensemble was 238,125 km^2 (97.5% of the total area inundated). However, when examining the hit rate in terms of the uncertainty in the ensemble, we find that the area that was inundated for the entire ensemble was 38,436 km^2 (15.7%), the inundated pixels correctly predicted by 2/3 of the ensemble was 109,965 km^2 (45.0%), while the area correctly inundated only in the maximum ensemble simulation was 89,723 km^2 (36.8%). The areas that were incorrectly classified as either nonflooded or flooded were 6,018 km^2 and 31,158 km^2, respectively. The false negatives corresponded to 2.2% of the area inundated either in the benchmark or reanalysis simulations. None of the ensemble members flooded those areas suggesting that the reanalysis precipitation was underestimated at the local scale resulting in inadequate flows in those areas. On the other hand, the false positives equaled 11.6% of the area inundated in the reanalysis simulations with the ensemble maximum

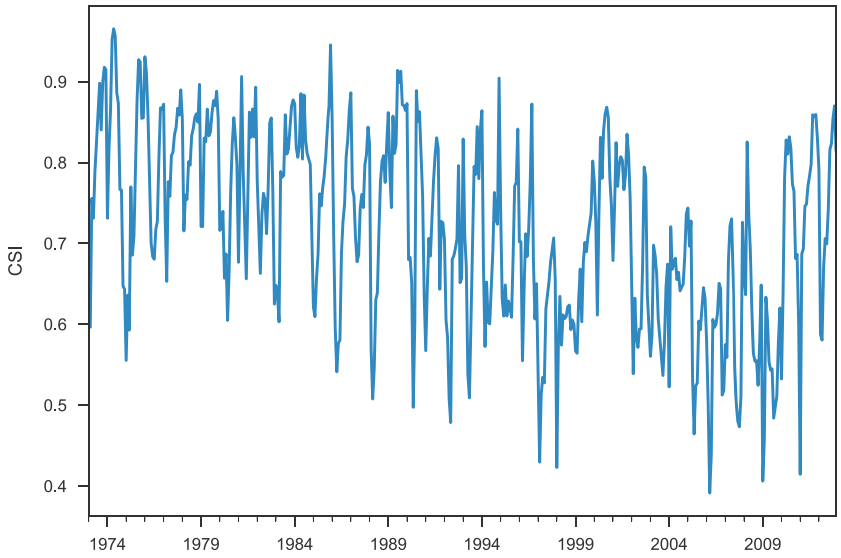

Figure 3. Time series of Critical Success Index (CSI) calculated from the flood depth ensemble and benchmark simulations.

contributing 9.5% (33.3% probability) while 2.1% of the total area was incorrectly predicted by 2/3 ensemble members (mean and maximum).

Most of the areas that were incorrectly inundated by the reanalysis ensemble were located in the Murray-Darling basin (region D), although false alarm pixels were apparent in all flooded regions. The flow regime for the Murray-Darling can be classified as intermittent with high daily flow variability as well as having stable baseflow near the coast (Kennard et al., 2010), although natural flows have been modified from human activities downstream. The intermittency in flows could lead to higher uncertainty and could partly explain the relatively large misclassification of inundated area for this region. On the other hand, the stable baseflow and a strong seasonal signal could have resulted in the correct classification of flood inundation. The reanalysis-derived inundated area in the regions in western and northern Australia (A and C in Figure 2) was correctly classified with reasonable confidence (2/3 probability), but was underestimated (misses) at the floodplains upstream, which is probably related to the degree of local detail in floodplain topography not captured by the SRTM DEM in those regions. The flow regime of both of these regions can be characterized as intermittent during the summer, dominated by zero-flow days or relatively constant flows with high predictability that could explain the lower uncertainty. Region B (Queensland) exhibits correct classification of inundated pixels for the ensemble mean and maximum traces, although there are some disjointed patterns of false alarms, which could be attributable to errors in the spatial variability of rainfall.

The map in Figure 2 does not have explicit timing information, but it does show that the reanalysis simulations do not inundate areas that were identified as flooded in the benchmark simulation. When combining this information with Figure 1, it is implied that the reanalysis simulations were not able to capture the spatial variability of the inflows leading to some areas not being inundated during the entire simulation period. It is apparent that the lower bound of the ensemble captures the inundated area to a small degree, mostly in areas adjacent to the river channels. On the other hand, there are large areas (especially in Queensland) when only the upper bound of the reanalysis ensemble results in true positives of maximum inundation extent.

Another approach to evaluate the estimation of flood inundation from the reanalysis simulations is by adapting a categorical verification metric such as the Critical Success Index (CSI). The definition of the CSI is given by

$$\mathrm{CSI} = \frac{A}{A + B + C} \tag{1}$$

where the A, B, and C terms are the number of pixels where an event is both predicted and observed, an event is predicted but not observed, and an event is observed but not predicted, respectively (Wu et al., 2012). The "event" is defined here as the flow depth value falling within the reanalysis ensemble for each pixel. Figure 3 shows a monthly time series of the CSI index calculated for the entire study area. The CSI metric allows the

evaluation of the reanalysis simulations in both space and time, by capturing both the magnitude and timing of predicted flow depth for each 1 km pixel. The prediction skill of flow depth appears to slightly decrease with time during the study period, although the median CSI is 0.72 suggesting a reasonably good agreement between the reanalysis ensemble and the benchmark simulation.

4. Conclusions

A model cascade that is driven by meteorological data and couples a hydrologic (VIC) and a high-resolution (1 km) hydrodynamic (LISFLOOD-FP) model was employed to evaluate whether reanalysis data sets could reproduce flood inundation over Australia during 1973–2012. An ensemble of simulations were performed by constructing an envelope of model traces around the mean reanalysis meteorology from the latter's associated uncertainty and were compared with a benchmark simulation that was driven by observed gauge inflows. The reanalysis ensemble was able to capture the inundated volume and area overall, although the uncertainty propagated from the meteorology resulted in a large ensemble spread. Moreover, the reanalysis-driven simulation produced a majority of true positives in terms of flood extent, although there were some problematic areas (misses and false alarms) suggesting that the inflow spatial and local topographic variability were not properly captured. Further analysis is needed to identify the factors affecting the performance of the reanalysis ensemble in terms of predicting flood inundation and also examine attribution of the resulting errors. Additionally, there are a number of approaches that could improve the reanalysis simulations including their merging with remotely sensed observations of inundated area or the correction of the meteorological data with other observational data sets (e.g., Tanoue et al., 2016). Nonetheless, our study presented and evaluated the application of a hydrologic/hydrodynamic model cascade over continental scales for long-term, continuous event simulations and demonstrated its feasibility for global implementation.

Acknowledgments

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration through the Advanced Information Systems Technology program (14AIST14-0074). Paul Bates was supported by a Royal Society Wolfson Research Merit Award. The authors would like to thank Augusto Geitrana and an anonymous reviewer for helping improve the manuscript. Data sets generated from this research are available at https://doi.org/10.6084/m9.figshare.5376430.

References

Andreadis, K. M., Schumann, G. J. P., & Pavelsky, T. (2013). A simple global river bankfull width and depth database. *Water Resour Research*, *49*(10), 7164–7168. https://doi.org/10.1002/wrcr.20440

Bates, P. D., Horritt, M. S., & Fewtrell, T. J. (2010). A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of Hydrology*, *387*(1–2), 33–45. https://doi.org/10.1016/j.jhydrol.2010.03.027

Brakenridge, G., Syvitski, J., Niebuhr, E., Overeem, I., Higgins, S., Kettner, A., & Prades, L. (2017). Design with nature: Causation and avoidance of catastrophic flooding, Myanmar. *Earth-Science Reviews*, *165*, 81–109. https://doi.org/10.1016/j.earscirev.2016.12.009

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., … Worley, S. J. (2011). The Twentieth Century Reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, *137*(654), 1–28. https://doi.org/10.1002/qj.776

Criss, R. E., & Winston, W. E. (2008). Do Nash values have value? Discussion and alternate proposals. *Hydrological Processes*, *22*(14), 2723–2725. https://doi.org/10.1002/hyp.7072

Dixon, M. J., & Tawn, J. A. (1995). A semi-parametric model for multivariate extreme values. *Statistics and Computing*, *5*(3), 215–225. https://doi.org/10.1007/BF00142663

Dolinar, E. K., Dong, X., & Xi, B. (2016). Evaluation and intercomparison of clouds, precipitation, and radiation budgets in recent reanalyses using satellite-surface observations. *Climate Dynamics*, *46*(7–8), 2123–2144. https://doi.org/10.1007/s00382-015-2693-z

Doocy, S., Daniels, A., Murray, S., & Kirsch, T. D. (2013). The human impact of floods: A historical review of events 1980–2009 and systematic literature review. *PLOS Currents*. https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a

Emerton, R., Cloke, H. L., Stephens, E. M., Zsoter, E., Woolnough, S. J., & Pappenberger, F. (2017). Complex picture for likelihood of ENSO-driven flood hazard. *Nature Communications*, *8*, 14796. https://doi.org/10.1038/ncomms14796

Essou, G. R. C., Sabarly, F., Lucas-Picher, P., Brissette, F., Poulin, A., Essou, G. R. C., … Poulin, A. (2016). Can precipitation and temperature from meteorological reanalyses be used for hydrological modeling. *Journal of Hydrometeorology*, *17*(7), 1929–1950. https://doi.org/10.1175/JHM-D-15-0138.1

Friedl, M. a., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, *114*(1), 168–182. https://doi.org/10.1016/j.rse.2009.08.016

Hagen, E., & Lu, X. X. (2011). Let us create flood hazard maps for developing countries. *Natural Hazards*, *58*(3), 841–843. https://doi.org/10.1007/s11069-011-9750-7

Hallegatte, S., Green, C., Nicholls, R. J., & Corfee-Morlot, J. (2013). Future flood losses in major coastal cities. *Nature Climate Change*, *3*(9), 802–806. https://doi.org/10.1038/nclimate1979

Hidalgo, H. G., Dettinger, M. D., & Cayan, D. R. (2008). Downscaling with constructed analogues: Daily precipitation and temperature fields over the United States (Tech. Rep.). Sacramento, CA: California Energy Commission. https://doi.org/CEC-500-2007-123

Hidalgo, H. G., Das, T., Dettinger, M. D., Cayan, D. R., Pierce, D. W., Barnett, T. P., … Nozawa, T. (2009). Detection and attribution of streamflow timing changes to climate change in the Western United States. *Journal of Climate*, *22*(13), 3838–3855. https://doi.org/10.1175/2009JCLI2470.1

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., … Stocker, E. F. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, *8*(1), 38–55. https://doi.org/10.1175/JHM560.1

Johnston, R. M., Barry, S. J., Bleys, E., Bui, E. N., Moran, K. C. J., Simon, D. A. P., … Grundy, M. (2003). ASRIS: The database. *Australian Journal of Soil Research*, *41*(6), 1021–1036. https://doi.org/10.1071/SR02033

Jones, D. A., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal*, *58*(4), 233–248.

Kennard, M. J., Pusey, B. J., Olden, J. D., MacKay, S. J., Stein, J. L., & Marsh, N. (2010). Classification of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology*, *55*(1), 171–193. https://doi.org/10.1111/j.1365-2427.2009.02307.x

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., … Takahashi, K. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Series II*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Lehner, B., & Doll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*(1–4), 1–22.

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, *99*(D7), 14,415–14,428. https://doi.org/10.1029/94JD00483

Lohmann, D., Rashke, E., Nijssen, B., & Lettenmaier, D. P. (1998). Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. *Hydrological Sciences Journal*, *43*(1), 131–141. https://doi.org/10.1080/02626669809492107

Maurer, E. P., & Hidalgo, H. G. (2008). Utility of daily vs. monthly large-scale climate data: An intercomparison of two statistical downscaling methods. *Hydrology and Earth System Sciences*, *12*(2), 551–563. https://doi.org/10.5194/hess-12-551-2008

Maurer, E. P., Hidalgo, H. G., Das, T., Dettinger, M. D., & Cayan, D. R. (2010). The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrology and Earth System Sciences*, *14*(6), 125–1138. https://doi.org/10.5194/hess-14-1125-2010

Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., … Ip, A. (2016). Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sensing of Environment*, *174*, 341–352. https://doi.org/10.1016/j.rse.2015.11.003

Neal, J., Schumann, G., & Bates, P. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resources Research*, *48*, 1–16. https://doi.org/10.1029/2012WR012514

Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D., & Wood, E. F. (2001). Predicting the discharge of global rivers. *Journal of Climate*, *14*(15), 3307–3323. https://doi.org/10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2

Paiva, R. C. D., Collischonn, W., & Buarque, D. C. (2013). Validation of a full hydrodynamic model for large-scale hydrologic modelling in the Amazon. *Hydrological Processes*, *27*(3), 333–346. https://doi.org/10.1002/hyp.8425

Pappenberger, F., Dutra, E., Wetterhall, F., & Cloke, H. L. (2012). Deriving global flood hazard maps of fluvial floods through a physical model cascade. *Hydrology and Earth System Sciences*, *16*(11), 4143–4156. https://doi.org/10.5194/hess-16-4143-2012

Sampson, C. C., Fewtrell, T. J., O'Loughlin, F., Pappenberger, F., Bates, P. B., Freer, J. E., & Cloke, H. L. (2014). The impact of uncertain precipitation data on insurance loss estimates using a flood catastrophe model. *Hydrology and Earth System Sciences*, *18*(6), 2305–2324. https://doi.org/10.5194/hess-18-2305-2014

Sampson, C. C., Smith, A. M., Bates, P. B., Neal, J. C., Alfieri, L., & Freer, J. E. (2015). A high-resolution global flood hazard model. *Water Resources Research*, *51*, 7358–7381. https://doi.org/10.1002/2015WR016954

Schumann, G. J. P., Neal, J. C., Voisin, N., Andreadis, K. M., Pappenberger, F., Phanthuwongpakdee, N., … Bates, P. D. (2013). A first large-scale flood inundation forecasting model. *Water Resources Research*, *49*, 6248–6257. https://doi.org/10.1002/wrcr.20521

Schumann, G. J.-P., Stampoulis, D., Smith, A. M., Sampson, C. C., Andreadis, K. M., Neal, J. C., & Bates, P. D. (2016). Rethinking flood hazard at the global scale. *Geophysical Research Letters*, *43*, 10,249–10,256. https://doi.org/10.1002/2016GL070260

Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, *19*(13), 3088–3111. https://doi.org/10.1175/JCLI3790.1

Tanoue, M., Hirabayashi, Y., & Ikeuchi, H. (2016). Global-scale river flood vulnerability in the last 50 years. *Scientific Reports*, *6*, 36021. https://doi.org/10.1038/srep36021

Trigg, M. A., Birch, C. E., Neal, J. C., Bates, P. D., Smith, A., Sampson, C. C., … Fewtrell, T. J. (2016). The credibility challenge for global fluvial flood risk analysis. *Environmental Research Letters*, *11*(9), 94014. https://doi.org/10.1088/1748-9326/11/9/094014

Ward, P., Jongman, B., Weiland, F. S., Bouwman, A., van Beek, R., Bierkens, M., … Winsemius, H. C. (2013). Assessing flood risk at the global scale: Model setup, results, and sensitivity. *Environmental Research Letters*, *8*(4), 44019. https://doi.org/10.1088/1748-9326/8/4/044019

Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., … Winsemius, H. C. (2015). Usefulness and limitations of global flood risk models. *Nature Climate Change*, *5*(8), 712–715. https://doi.org/10.1038/nclimate2742

Whelan, J., & Frederiksen, J. S. (2017). Dynamics of the perfect storms: La Niña and Australia's extreme rainfall and floods of 1974 and 2011. *Climate Dynamics*, *48*(11-12), 3935–3948. https://doi.org/10.1007/s00382-016-3312-3

Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J., & Bouwman, A. (2013). A framework for global river flood risk assessments. *Hydrology and Earth System Sciences*, *17*(5), 1871–1892. https://doi.org/10.5194/hess-17-1871-2013

Wu, H., Adler, R. F., Hong, Y., Tian, Y., & Policelli, F. (2012). Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. *Journal of Hydrometeorology*, *13*(4), 1268–1284. https://doi.org/10.1175/JHM-D-11-087.1

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, *47*, 1–21. https://doi.org/10.1029/2010WR009726

Zsótér, E., Pappenberger, F., Smith, P., Emerton, R. E., Dutra, E., Wetterhall, F., … Balsamo, G. (2016). Building a multimodel flood prediction system with the TIGGE archive. *Journal of Hydrometeorology*, *17*(11), 2923–2940. https://doi.org/10.1175/JHM-D-15-0130.1